

Project Proposal

MIS 545 Summer 2020

Ryan Straight

Abstract

Assignment submitted to MIS 545 Summer 2020 on 2020-07-30. R packages and versions used in this document: R (Version 4.0.0; R Core Team, 2020) and the R-package *papaja* (Version 0.1.0.9942; Aust & Barth, 2020). Any *LaTeX* is generated using the [LaTeX Equation Editor](#).

Word count: 362

Project Proposal

Given the timeliness of COVID-19¹ and the considerable amount of granular data available about confirmed cases, school and business closings, travel restrictions, et cetera, this will be the main focus of the project. The main goal of this project is to determine what, if any, predictors for success and financial loss mitigation exist within the data available.

Business Value

The potential financial hardship caused by school closings and resultant educational delays on the future workforce cannot be overstated. The amount of future earnings decreased by even a scant four months of higher educational closures might not seem detrimental but “the loss in marginal future earnings would be 2.5 percent per year over a student’s working life” (Psacharopoulos, Patrinos, Collis, & Vegas, 2020). Applied to the 76 million students and presuming a working-life of 45 years at an average annual income of \$53,490, the Brookings Institute estimates an annual loss of \$1,337 per student, resulting in a staggering \$2.5 trillion for those four months of lost education (2020).

Data Source

There is no shortage of sources for COVID-19 data. In an effort to ensure the data used in this project is up-to-date at time of publishing and will be consistently formatted throughout, the data will be pulled using the COVID19 package (Guidotti & Ardia, 2020). This data is combined

¹ COVID-19 is the illness caused by the SARS-CoV2 virus, not the virus, itself.

with the “COVID-19 Lockdown dates by country” data set (jczyag, 2020) and the list of states’ governors and political party affiliation (Civil Services, 2020).

Data Mining Techniques

While specific techniques are presumptive at this point given the draft stage of the current project, the identified statistical tests and techniques are clustering descriptive analysis, decision trees, and a naïve Bayesian classification. As the data is analyzed and explore further this may change, however.

References

- Aust, F., & Barth, M. (2020). *papaja: Create APA manuscripts with R Markdown*. Retrieved from <https://github.com/crsh/papaja>
- Civil Services. (2020). United States Governors. <https://civilserviceusa.github.io/us-governors/>.
<https://civilserviceusa.github.io/us-governors/>.
- Guidotti, E., & Ardia, D. (2020). COVID-19 Data Hub. *Journal of Open Source Software*, 5(51), 2376. <https://doi.org/10.21105/joss.02376>
- jcyzag. (2020). COVID-19 Lockdown dates by country. Retrieved from <https://kaggle.com/jcyzag/covid19-lockdown-dates-by-country>
- Psacharopoulos, G., Patrinos, H., Collis, V., & Vegas, E. (2020). The COVID-19 cost of school closures. *Brookings*.
- R Core Team. (2020). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <https://www.R-project.org/>